

UCLA

UCLA Previously Published Works

Title

The commonness of rarity: Global and future distribution of rarity across land plants.

Permalink

<https://escholarship.org/uc/item/8q35827b>

Journal

Science advances, 5(11)

ISSN

2375-2548

Authors

Enquist, Brian J
Feng, Xiao
Boyle, Brad
et al.

Publication Date

2019-11-27

DOI

10.1126/sciadv.aaz0414

Peer reviewed

ECOLOGY

The commonness of rarity: Global and future distribution of rarity across land plants

Brian J. Enquist^{1,2*}, Xiao Feng³, Brad Boyle¹, Brian Maitner¹, Erica A. Newman^{1,3}, Peter Møller Jørgensen⁴, Patrick R. Roehrdanz⁵, Barbara M. Thiers⁶, Joseph R. Burger³, Richard T. Corlett⁷, Thomas L. P. Couvreur⁸, Gilles Dauby⁹, John C. Donoghue¹⁰, Wendy Foden¹¹, Jon C. Lovett^{12,13}, Pablo A. Marquet^{2,14,15}, Cory Merow¹⁶, Guy Midgley¹⁷, Naia Morueta-Holme¹⁸, Danilo M. Neves¹⁹, Ary T. Oliveira-Filho¹⁹, Nathan J. B. Kraft²⁰, Daniel S. Park²¹, Robert K. Peet²², Michiel Pillet¹, Josep M. Serra-Diaz²³, Brody Sandel²⁴, Mark Schildhauer²⁵, Irena Šímová^{26,27}, Cyrille Violle²⁸, Jan J. Wieringa²⁹, Susan K. Wiser³⁰, Lee Hannah⁵, Jens-Christian Svenning³¹, Brian J. McGill³²

Copyright © 2019
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

A key feature of life's diversity is that some species are common but many more are rare. Nonetheless, at global scales, we do not know what fraction of biodiversity consists of rare species. Here, we present the largest compilation of global plant diversity to quantify the fraction of Earth's plant biodiversity that are rare. A large fraction, ~36.5% of Earth's ~435,000 plant species, are exceedingly rare. Sampling biases and prominent models, such as neutral theory and the k-niche model, cannot account for the observed prevalence of rarity. Our results indicate that (i) climatically more stable regions have harbored rare species and hence a large fraction of Earth's plant species via reduced extinction risk but that (ii) climate change and human land use are now disproportionately impacting rare species. Estimates of global species abundance distributions have important implications for risk assessments and conservation planning in this era of rapid global change.

INTRODUCTION

Why some species are common and others are rare has intrigued ecologists (1, 2), at least, since Darwin (3). Rare species are orders of magnitude more likely to go extinct (4, 5), making it puzzling how so many rare species can be maintained (6). Understanding rarity and the maintenance of rare species is also central to conservation biology [e.g., (7)] and to understanding current and future changes in biodiversity due to global change (8). Despite this importance, we know unexpectedly little about the causes of commonness and rarity and their maintenance at a global scale (9, 10).

Most quantifications of species abundance use abundances in local communities because estimates of global taxon abundance are difficult to obtain. However, there are two major limitations to focusing solely on local abundance. First, most species tend to be simultaneously common in a few parts of their ranges and rare in most of their ranges (11, 12), making estimates of local abundance a noisy and a poor measure of how truly rare a species is globally. Second, at a global

scale, a measure of rarity results from a combination of the average local abundance and the number of sites occupied throughout the species geographic range. Local species abundance and species occupancy across the geographic range tend to be correlated (12–14), so locally rare species tend to also show up in only a few local communities. This makes it likely that estimates of global abundance will be more skewed to the rare, but this has rarely been tested (15). A global estimate of rarity can therefore minimize the potential problems associated with assessing whether a species is rare. Fortunately, with the rapid development of biodiversity databases and networks in the past decade, it is becoming increasingly possible to quantify continental and global patterns of biodiversity and test competing models for the origin and maintenance of these patterns at a global scale (16).

Here, we use a global botanical database of unprecedented coverage to (i) assess global patterns of plant rarity, (ii) test several proposed hypotheses underlying the generation and persistence of rare species, (iii) identify regions that harbor hotspots of rare species and explore

¹Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA. ²Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA. ³Institute of the Environment, University of Arizona, Tucson, AZ 85721, USA. ⁴Missouri Botanical Garden, St. Louis, MO 63110, USA. ⁵Betty and Gordon Moore Center for Science, Conservation International, 2011 Crystal Dr., Arlington, VA 22202, USA. ⁶New York Botanical Garden, 2900 Southern Blvd., Bronx, NY 10348, USA. ⁷Centre for Integrative Conservation, Xishuangbanna Tropical Botanical Garden and Center of Conservation Biology, Core Botanical Gardens, Chinese Academy of Sciences, Menglung, Yunnan, China. ⁸DIADÉ, IRD, Université Montpellier, Montpellier, France. ⁹AMAP, IRD, CIRAD, CNRS, INRA, Université Montpellier, Montpellier, France. ¹⁰Rincon Consultants Inc., Ventura, CA 93003, USA. ¹¹Cape Research Centre, South African National Parks, Tokai, 7947 Cape Town, South Africa. ¹²School of Geography, University of Leeds, Leeds, UK. ¹³Royal Botanic Gardens, Kew, Richmond, Surrey, UK. ¹⁴Departamento de Ecología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, CP 8331150 Santiago, Chile. ¹⁵Instituto de Ecología y Biodiversidad (IEB), Laboratorio Internacional de Cambio Global and Centro de Cambio Global UC, Chile. ¹⁶Department of Ecology and Evolutionary Biology, University of Connecticut, CT 06269, USA. ¹⁷Department of Botany and Zoology, Stellenbosch University, Stellenbosch, South Africa. ¹⁸Center for Macroecology, Evolution and University of Copenhagen, Universitetsparken 15, Building 3, DK-2100 Copenhagen Ø, Denmark. ¹⁹Department of Botany, Federal University of Minas Gerais, Belo Horizonte 31270-901, Minas Gerais, Brazil. ²⁰Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA 90095, USA. ²¹Department of Organismic and Evolutionary Biology, Harvard University, MA 02138, USA. ²²Department of Biology, University of North Carolina, NC 27599, USA. ²³Université de Lorraine, AgroParisTech, INRA, Silva, 54000 Nancy, France. ²⁴Department of Biology, Santa Clara University, Santa Clara, CA 95053, USA. ²⁵National Center for Ecological Analysis and Synthesis, Santa Barbara, CA 93101, USA. ²⁶Centre for Theoretical Study, Charles University, Prague 1, Czech Republic. ²⁷Department of Ecology, Faculty of Sciences, Charles University, Czech Republic. ²⁸Université Montpellier, CNRS, EPHE, IRD, Université Paul Valéry Montpellier 3, Montpellier, France. ²⁹Naturalis Biodiversity Center, Darwinweg 2, Leiden, Netherlands. ³⁰Manaaki Whenua—Landcare Research, Lincoln, New Zealand. ³¹Center for Biodiversity Dynamics in a Changing World (BIOCHANGE) and Section for Ecoinformatics and Biodiversity, Department of Bioscience, Aarhus University, Ny Munkegade 114, DK-8000 Aarhus C, Denmark. ³²School of Biology and Ecology and Senator George J. Mitchell Center of Sustainability Solutions, University of Maine, Orono, ME 04469, USA.

*Corresponding author. Email: benquist@email.arizona.edu

the drivers of these spatial patterns, and (iv) assess how current patterns of human impact and future climate change scenarios may affect plant diversity via impacts on rare species. In the past, quantification of global patterns of abundance and rarity has been hampered by the many limitations of global biodiversity data. These issues have made the use of these data in comprehensive biodiversity analyses difficult (17, 18). Here, we take a novel approach that overcomes many of these limitations. For all known land plants (Embryophyta), we have compiled a global database of standardized botanical observation records—the integrated Botanical Information and Ecology Network (BIEN) [Fig. 1, BIEN v4.1; <http://bien.nceas.ucsb.edu/bien/>; see the Supplementary Materials; (19)]. The BIEN data are mainly composed of herbarium collections, ecological plots and surveys, and trait observations. Together, these data constitute more than 200 million observations of plant species occurrences. Assembling these data involves overcoming numerous challenges of taxonomy, data quality, data exchange, provenance, interoperability, and scaling (Fig. 1) (20). After correcting misspelled or synonymous taxon names and removing records with invalid or suspect geocoordinates, incomplete or unresolvable taxon names, and observations of non-native species and cultivated plants, the final dataset consists of 34,902,348 observation records of 434,934 land plant species from herbarium and ecological plot data (see Fig. 1 and the Supplementary Materials for details of data cleaning and validation).

We quantified the distribution of global abundance for all land plant species (hereafter, plant species) on Earth using a metric of global relative abundance, the total number of unique observations of a species ever recorded in global databases. The distribution of the total number of global observations per species [the global species

abundance distribution (gSAD)] is an estimate of global abundance and is still a sample, as a count of all individuals on the planet is impossible. Nonetheless, quantifying the functional form of gSADs has a substantial practical advantage over other estimates of abundance. First, we can combine data from different datasets including plots and surveys, and herbarium specimens to increase sampling coverage. These datasets all share the common attribute of observing an individual of a given species in a given location and time. Second, comparing and integrating estimates of gSADs from different datasets (e.g., plots versus herbarium specimens) provide a way to assess potential biases in estimating species global abundance. For example, gSADs can be estimated by compiling only plot or ecological survey data. In plot data, a global estimate of species abundance is quantified directly, as each individual of that species is summed within and across plots. As we discuss, our approach is less biased than local plot-based abundance data that samples only a tiny fraction of Earth's surface.

Traditionally, measures of rarity have been based on a multi-dimensional concept. For example, Rabinowitz (21) identified three major axes on which a species can be common or rare: local abundance, extent of the geographic range, and habitat specificity. Although conceptually these three dimensions are independent, they are often strongly positively correlated (22). Four of the five criteria the International Union for Conservation of Nature uses to evaluate extinction risk for their Red List (23) directly involve measurement of rarity via absolute levels of or declines in abundance and geographic distribution, while the fifth involves computer simulations, which are likely to incorporate population size and range size as well. These criteria all point to the importance of measuring rarity at

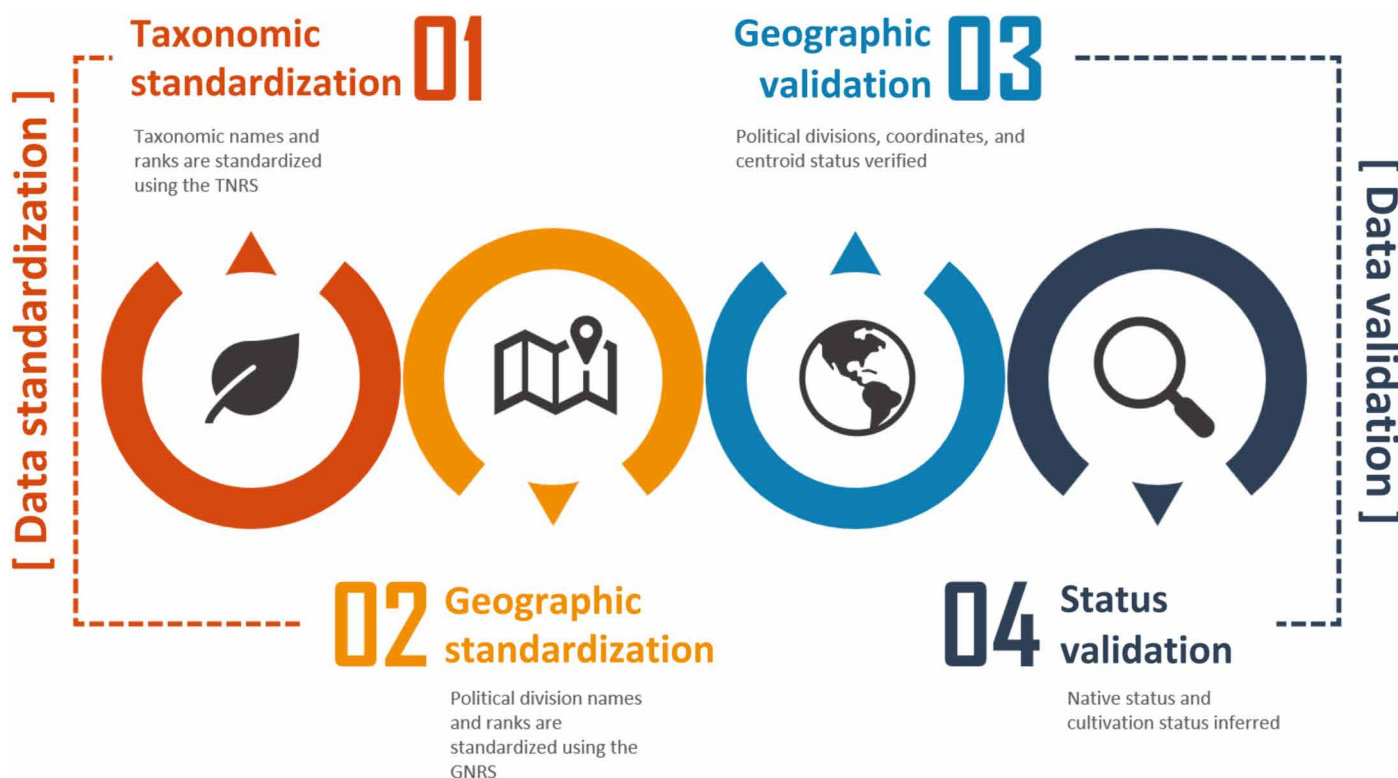


Fig. 1. Computational workflow for creating gSADs. TNRS, Taxonomic Name Resolution Service; GNRS, Geographic Name Resolution Service.

global scale (in contrast to local rarity). A species may be globally rare because it has few individuals at many sites or many individuals at few sites.

RESULTS

We generated three gSAD distributions based on summing individual observations of species across all ecological plots, by summing all observations across all other botanical observation records, and by summing the non-plot observation records found within 100 km distance from each ecological plot. Our analyses reveal that a large fraction of the plant species on Earth are rare (Figs. 2 and 3). Analyzing the distribution of the number of observations per species reveals that the global-scale distribution is highly skewed and lacking a central tendency (i.e., the mode of the gSAD is at $N = 1$; Fig. 2). The total number of land plant species is ~435,000 (the number of species before geovalidation based on 66,334,188 observation records), a large fraction of these species, 36.5% or 158,535, are rare, with just five observations or fewer, while 28.3% or 123,149 have just three observations or fewer. The large number of rare species is consistent with past claims that when biodiversity observations are compiled at increasingly larger spatial (15) and temporal scales (24), rare species should comprise an increasing majority of species.

Global species abundance distribution

We tested several long-standing hypotheses concerning the processes creating and maintaining large-scale patterns of commonness and rarity. Specifically, we assessed whether the number of observed rare species follows predictions from biodiversity theory by comparing several proposed statistical distributions for the gSAD. First, we assessed two contrasting sets of predictions for the distribution of commonness and rarity of species (Fig. 2). Spe-

cifically, at increasingly larger geographic scales, both the unified neutral theory of biogeography (UNTB) (25) and the k -niche model (26) predict that the gSAD will converge on Fisher's log-series distribution (27)

$$\hat{f} = \frac{\alpha x^n}{n} \quad (1)$$

where \hat{f} is the expected number of species, n is the total number of observations per species, α is the diversity parameter, and x is a nuisance parameter that is defined by α and the total number of individuals sampled, N , $x = N/(N - \alpha)$. The UNTB further makes two predictions: (i) At increasingly large spatial scales (such as continental and global scales), the Fisher's log-series distribution will also increasingly converge to approximate a "power law" (or a Pareto distribution) over most of the range of the distribution (see Fig. 2A) (28), where

$$\hat{f} = \frac{\beta - 1}{n_0} \left(\frac{n}{n_0} \right)^{-\beta} \quad (2)$$

where (ii) the value of β , the scaling exponent or slope on a log-log plot, will equal -1.0 . For the continuous Pareto or power law distribution, n_0 is the minimum scale of the distribution, and β is the scaling exponent (29). For the BIEN data, the minimum number of observations for a species is 1, so it was set at 1.

The UNTB predicts that the gSAD (called the regional pool in neutral theory) will follow a log-series distribution. Pueyo (28) notes that the log-series distribution consists of two parts multiplied together: a Pareto distribution with exponent $\beta = -1$ that is the result of neutral dynamics and an exponential "bend" that takes effect at very high abundances due to the finite size assumption. Pueyo (28) also suggests a generalization of the Pareto and log series that incorporates a Pareto where the exponent β is allowed to vary

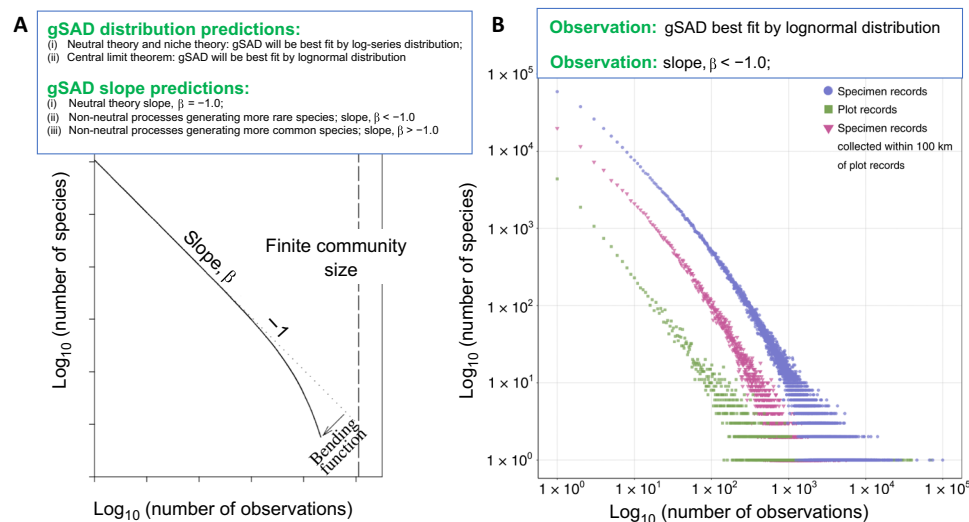


Fig. 2. The gSAD for all plant species. (A) Schematic illustration of the predicted gSAD based on expectations from theory (see main text) (28). In the inset, we list several differing predictions for the shape of the gSAD. (B) Two estimates of the gSAD for all land plant species. The first distribution (green) is the observed number of observations per species for all species found in ecological plots. Each data point represents the total number of individuals observed for a given species. The second distribution (red) is all botanical specimens collected within 100 km of each plot. The third distribution (light purple) is all botanical specimens per species. Each distribution is strongly modal at the lowest abundance, showing that most species have only been observed a very small number of times and only a few species are common. The distributions are shown on \log_{10} -transformed axes. Comparing the shape of the distributions of the competing fits of differing proposed gSAD distributions allows us to test differing hypotheses for the origin of the gSAD.

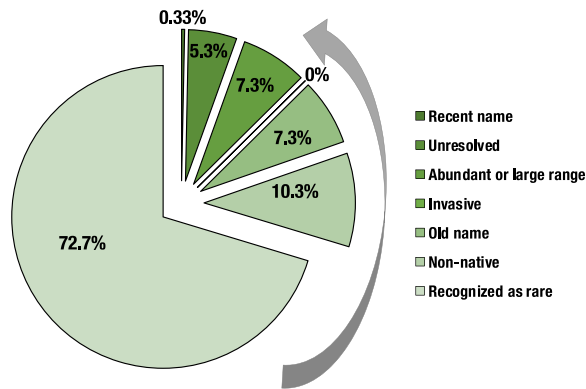


Fig. 3. Does using the number of observations in botanical datasets provide a reliable measure of rarity? Assessments of rarity by taxonomic specialists at the Missouri Botanical Garden and the New York Botanical Garden for a random sample of 300 species with three observations or fewer in the BIEN database. Most species (72.7%) identified as “rare” based on the number of unique occurrences within the BIEN database are also recognized as rare by experts. Approximately 7.3% of these species appear to be incorrectly characterized as rare, as they are recognized by experts as abundant or having large ranges. The apparent scarcity of approximately 7.5% of these taxa may reflect recent taxonomic splits or old names no longer used. Moreover, 10.3% are non-native species (which may or may not be rare). In sum, we estimate that between 72 and 90% of plant taxa (recognized as rare + recent name + unresolved + old name) identified by BIEN as being rare would be recognized as rare by other measures.

combined with an exponential finite size term, which we call here “Pareto with exponential finite adjustment.” Thus, testing whether the gSAD is best fit by a log series (where $\beta = -1$), a Pareto distribution (where β is allowed to vary), or a Pareto with exponential finite adjustment (where β is also allowed to vary) provides a test of neutral dynamics. In sum, both the UNTB and k-niche model predict that the log-series distribution will best fit gSADs, but at large geographic scales, this distribution will also converge to a Pareto distribution. Thus, fitting the Pareto or the Pareto with exponential finite adjustment provides a simultaneous test of whether neutral or niche dynamics are consistent with the data (28). A poor fit or a value of $\beta \neq 1.0$ rejects neutral theory. A poor fit of the Pareto regardless of the value of β further rejects the k-niche model (28). In addition, the value of β is then a useful ecological and evolutionary indicator of whether Earth has more rare species ($\beta < -1$; the slope of the function is steeper) or fewer ($\beta > -1$; the slope of the function is flatter) rare species than expected under zero-sum neutral evolutionary dynamics (28, 30).

In contrast to the predictions from the UNTB and k-niche model, the central limit theorem (CLT) predicts that gSADs will be characterized by a lognormal distribution. If the abundance of a species is the result of several multiplicative processes acting together (31), then lognormal distributions are expected. Because of the CLT, a lognormal distribution is expected any time many variables interact multiplicatively to influence abundance, such as many differing biotic and abiotic factors [see references in (32)]. Common processes in ecology and evolution are known to interact multiplicatively to influence species abundance (see Supplementary Document) (32). One context in which random variables are multiplied (yielding a lognormal) is consecutive annual population growth rates, although the applicability of this across species (i.e., to generate SADs) is

controversial (33), relying on subtle philosophical interpretations of exchangeability. Some authors such as May (34) and MacArthur (35) say it can, while others such as Pielou (36) (see page 48) say it cannot produce a lognormal. This debate, however, is a red herring because many other biological processes in ecology and evolution also interact multiplicatively and can influence variation in inter-specific abundance. For example processes that lead to niche partitioning, stochastic density-dependent differential equation models (37), models of rates of fixation of favorable alleles (35), or hurdle models (15) can generate lognormal SADs. Note that in the case of discrete abundances sampled from a continuous lognormal, we have a Poisson lognormal (38).

Next, we fit several additional models and statistical distributions that have been proposed to describe the distribution of commonness and rarity [see (39, 40) and the Supplementary Materials]. Using maximum likelihood estimations (MLEs), we fit each distribution to three ways to assess empirical gSAD: (i) for all of the species observation records within the BIEN database, (ii) for all species recorded only from ecological plots, and (iii) for all specimens found within 100 km around each ecological plot. Comparing the goodness of fit of various models for each of these gSADs allows us to compare potential sampling biases in botanical data.

The best model varied with which measure of goodness of fit was used, as well as with the dataset used (Tables 1 and 2 and tables S1 and S2). However, in general, the truncated Pareto (i.e., a modified Pareto distribution that adds an additional parameter to allow the right tail to drop down because of finite sample size (28)) and the Poisson lognormal (41) both fit well. These models have strong skew on a log scale, indicative of many rare species. All three models (at the estimated parameter values) show the mode at species with one individual. The log series, while also showing a mode at one individual in a log plot, markedly underestimates the number of extremely rare species, and the remaining models fit the distribution even more poorly and have an interior mode, incorrectly predicting that the most common abundances will be intermediate.

The UNTB predicts that log-series distribution will be approximated by the fit of the Pareto power distribution, with an exponent, $\beta = -1.0$ (28). However, our fit of the log-series distribution shows that it was not the best fit and the fitted scaling exponent is steeper than -1.0 (β MLE = -1.41 for all of the BIEN observations and β MLE = -1.43 for the observations from ecological plot data; Fig. 2). Thus, a Pareto power distribution needs an exponent less than -1 to generate the number of rare species actually observed.

Together, these results underscore that, at continental to global scales, only a few species abundance distribution models are capable of producing sufficient numbers of rare species to match the observed data and that neutral dynamics under the UNTB is not one of them. The observed value of β for embryophytes is similar to what has been reported for an extensive dataset for other taxa including animals and marine phytoplankton (28), suggesting that the shape of the SAD at increasingly larger spatial scales may converge to a similar distribution across disparate taxa. In sum, the Poisson lognormal is best fit, the Pareto exponent is markedly steeper than -1.0 , and the Pareto distribution is the second best fit on two of the three metrics.

Assessment of sampling or taxonomic bias

Next, an obvious question is whether the observed number of rare species is the result of sampling or taxonomic bias. Data from herbarium records are known to exhibit biases in collection and sampling

Table 1. Three different measures of goodness of fit (r^2 or percentage of variance explained in the cumulative distribution function, χ^2 on \log_2 bins, and Akaike's information criterion) are shown for six different species abundance models [see (40)]. All distributions shown have two parameters except the log-series and power distributions, which have one. Distributions were fitted for the number of observations per species across all species found (i) within ecological plots only and (ii) across all datasets within the BIEN database. Sampling species found only in plots standardizes for sampling influences, as all individuals within ecological plots are sampled and identified to species. Thus, the species abundance distribution from ecological plots is expected to more accurately describe the species abundance distribution. As predicted by the CLT, the Poisson lognormal distribution provides the best fit to both gSADs. Nonetheless, Pareto and truncated Pareto also all fit well. The log-series distribution, predicted by the k-niche model and neutral theory, falls behind these distributions across the different goodness-of-fit measures. AIC, Akaike's information criterion; CDF, cumulative distribution function.

Model	Plot data only				All data			
	CDF r^2	$\chi^2 \log_2$	AIC	Δ AIC	CDF r^2	$\chi^2 \log_2$	AIC	Δ AIC
Zipf-Mandelbrot	0.929	54,188	139,822	25,848	0.447	73,884,947	7,402,206	330,9517
Weibull	0.999	1.6×10^{10}	127,111	13,137	0.999	3.01×10^{10}	4,269,287	176,598
Log series	0.991	1.57×10^{13}	120,082	6109	0.999	5.08×10^{13}	4,119,057	26,368
Pareto	0.999	5.69×10^{13}	115,244	1270	0.999	1.46×10^{13}	4,110,900	18,211
Poisson lognormal	0.999	490	113,974	0	0.999	2966	4,092,689	0
Pareto with finite sample exponential adjustment	0.999	563	114,096	122	0.998	100,558	4,203,550	110,861

Table 2. Parameter fits for each of the fitted statistical distributions. The estimated slope values, β , of the gSAD are given in bold by fits of the Pareto and Truncated Pareto distributions. Note that the estimated slope values differ from -1.0 expected from the unified neutral theory of biodiversity. Instead, the observed fitted slope, β , is steeper than expected from neutral theory (with fitted exponents more negative than -1.0). The steeper exponents indicate that of all of the observed plant species on Earth, proportionally more of them are rare and that there are more rare species than expected by demographic and evolutionary neutral processes. Thus, the processes creating and maintaining rare species on Earth generate proportionally more rare species.

Model	Plot data	All data
Zipf-Mandelbrot, b	13.3	1186.7
Zipf-Mandelbrot, c	1.4	1.2
Log series, c	0.9	0.9
Pareto fitted exponent, β	-1.4	-1.3
Weibull scale	18.1	40.6
Weibull shape	0.4	0.5
Poisson lognormal, m	4.07×10^{-8}	1.7
Poisson lognormal, s	2.9	2.6
Pareto with finite sample exponential adjustment (28) fitted Pareto exponent, β	-1.3	-1.1
Pareto with finite sample exponential adjustment: Exponential parameter, Ω	0.1	0.1

(17, 18). However, do these biases influence our identification of whether a species is rare or not?

We conducted two tests: First, in Fig. 2, we compared the distributions of global abundance in (i) the total BIEN database (including

plot surveys and herbarium records), (ii) only the plot datasets, and (iii) the subset of herbarium records that reflect the same geographic distribution as the survey data (e.g., all records within 100 km of any plot location) (Fig. 2). Ecological plots and surveys, in contrast to herbarium data, contain less sampling bias, as a robust effort is made to ensure all individuals within the sampling design are surveyed within a given area. In many cases, repeated visits ensure accurate identification to species. Thus, assessing whether the gSAD from plot data is different from the gSAD from all botanical observations based on sampling herbarium data at the globe or around plots enables us to assess potential bias and sampling effectiveness. As discussed below, both empirical gSADs are described by similar statistical distributions (e.g., the shape of gSADs in Fig. 2B are similar to each other and to Fig. 2A; Tables 1 and 2 and tables S1 and S2), indicating that sampling issues do not greatly influence conclusions regarding gSADs.

Next, to further assess whether rare species are truly rare or artifactual, we randomly sampled 300 rare species with three observations or fewer from the Americas. The Americas were chosen because our taxonomic expertise was focused on these two continents. For each species selected, we consulted taxonomic experts at the Missouri Botanical Garden and the New York Botanical Garden to sort each species into several classifications (Fig. 3 and see the Supplementary Materials). Taxonomic experts largely confirm that the majority of rare species identified by BIEN are rare, with only 7.3% that were clearly erroneous and recognized as abundant or large-ranged species. We conclude that our results are not driven by taxonomic and sampling biases.

Our results from Fig. 3 allow us to estimate the total number of native land plant species currently observed across the globe with estimates for taxonomic uncertainty. After correcting and standardizing data, we estimate that the total number of extant embryophyte (land plant) species on Earth is between ~358,000 and ~435,000. The lower limit stems from subtracting 17.8% from the total [10.3% from the remaining presence of naturalized non-native species +7.5%

caused by the inflation of names due to “old names” (basionyms) not yet corrected for by taxonomic data cleaning; see Fig. 3]. Our estimates are consistent with previous estimates of the total number of embryophytes in the world of approximately 391,000 [see (42)] or 403,911 (43) (see the Supplementary Materials). However, now, we can quantify that ~36% of these species are highly rare with very little distributional information for each species. In sum, our results from Fig. 2 show that rarity is commonplace across the land plants. Little botanical information exists across the world’s herbaria and ecological collections for between 11.2 and 13.6% (species with one observation) or between 30.0 and 36.5% (species with fewer than or equal to five total observations) of all vascular plant species.

“Hotspots” of rare species

To identify the regions that harbor hotspots of rare species, in Fig. 4, we mapped the locations of rare species across the world. We controlled for variation in sampling effort by calculating both the Menhinick and Margalef indices (see Materials and Methods). Plotting the sampling-corrected number of rare species reveals several patterns. Rare species cluster in the Americas in (i) mountainous regions (particularly along the thin strip along the western flank of the Andean Mountains, Central America, and the southern Sierra Madre of Mexico), (ii) the Guiana shield in northern South America, and (iii) relatively small climatic regions that are strongly distinct from surrounding areas (the Atlantic Forest or Mata Atlântica in Brazil, the southern region of the California Floristic Province, and the Caribbean); in Africa in (iv) the Cape Floristic Region of South Africa, (v) mountainous regions of Madagascar, (vi) the coastal mountains of Cameroon, and (vii) the Ethiopian highlands and the Somali peninsula; and in Asia in (viii) southwestern China and the border regions of Myanmar, Laos, and Thailand, (ix) Malaysia, (x) New Guinea, and (xi) the mountainous strip from Iran through Turkey. In Europe, there are several regions of notably high diversity of rarity in and around (xii) the Mediterranean, including the Pyrenees and Caucasus.

There is a relative dearth of rare species throughout the Amazon basin, confirming past claims that the Amazon flora consists largely of widespread and relatively abundant species (44). The areas identified by our methods show some overlap with areas independently identified as biodiversity hotspots (45) (e.g., Mesoamerican highlands, the Andes, Southeast Asia, and New Guinea) but differ in other areas.

Drivers of the spatial distribution of rarity

To assess the drivers of the spatial distribution of rarity, we conducted ordinary least squares (OLS) linear regression and simultaneously autoregressive (SAR) models to analyze the relationship between rarity index and environmental variables, including present climate, glacial-interglacial climatic velocity or instability of climate, and topography. Our OLS models showed that all the variables (annual mean temperature, annual precipitation, temperature seasonality, precipitation seasonality, temperature velocity, precipitation velocity, elevation, and heterogeneity of elevation) have significant relationships with both the Menhinick rarity index (tables S3 to S5 and fig. S2) and the Margalef rarity index (tables S6 to S8 and fig. S3), with the largest effects from temperature velocity and heterogeneity of elevation. In comparing the group models [present climate (annual mean temperature, annual precipitation, temperature seasonality, and precipitation seasonality), stability

of climate (temperature and precipitation velocity), and topography (elevation and heterogeneity of elevation)], the model with instability of climate tended to outperform models with current climate and topography, while the full model showed the lowest Akaike’s information criterion (AIC). The exhaustively selected model did not include elevation as a predictor, although it had minor differences in model performance compared with the full model.

A Moran’s *I* test showed high spatial autocorrelation in the residuals of the OLS models, while we found no significant spatial autocorrelation in the residuals of the SAR models (tables S3 to S8). The coefficients of the SAR models were generally similar to those from OLS models, with the exceptions that signs of annual mean temperature, precipitation seasonality, and precipitation velocity switched from positive to negative in the SAR model. Temperature velocity remains the largest negative effect, and heterogeneity of elevation remains the largest positive effect in the SAR models (see figs. S2 and S3 and tables S3 to S8). Models incorporating climate stability and topography outperformed the model with current climate, while the full model remains the best-performing SAR model. The modeling results based on Menhinick and Margalef rarity index showed comparable results (tables S3 to S8 and figs. S2 and S3).

To summarize, areas that contain a higher number of rare species have had a more stable climate. The best predictor of plant rarity is the historical temperature velocity. Climate velocity describes climate instability with ecologically relevant units (distance/time; see discussion in Supplementary Document). In addition, mountainous area, as measured by the SD of elevation, is also a predictor with positive effect (tables S3 to S8 and figs. S2 and S3). Adding short-term annual variation (annual seasonality) in temperature and precipitation and mountainous conditions in addition to climate velocity does improve the explanation of the current spatial distribution of rarity (e.g., the proportion of variation explained, R^2 , increased to 0.193 for the OLS model and to 0.518 for the SAR model of Menhinick rarity index but less so for Margalef rarity index, 0.176 for the OLS model and 0.263 for the SAR model; tables S3 to S8). Together, these results are consistent with previous results [see (46, 47) and references therein], indicating that increased rates of climate change velocity negatively affect the retention of rare species, presumably because of increased rates of extinction during times of rapid climate change.

The overlaps between future climate velocity and human footprint and rarity indices

Our environment is facing rapid human changes at the global scale, so we quantified the intensity of human impact on the area with rare species (48). Regions with rare species are currently characterized by higher human impact and will experience faster rates of future climate change under representative concentration pathway 8.5 (RCP8.5) (Fig. 5). Areas with rare species have human footprint values of 8.5 ± 5.8 , which is ~1.6 times higher ($P < 0.001$, Wilcoxon test) than that of the globe on average (5.2 ± 5.8). Furthermore, on average, areas with rare species are predicted to experience ~200 (± 58) times greater rates of temperature velocity than those same areas experienced historically in terms of the overall glacial-interglacial climate shift across the past 21,000 years [from the last glacial maximum (LGM) to the present]. The ratio between future temperature velocity and this long-term overall historical temperature velocity is ~1.2 times greater ($P < 0.001$, Wilcoxon test) for areas with rare species than the globe will experience on average (170 ± 77) (Fig. 5).

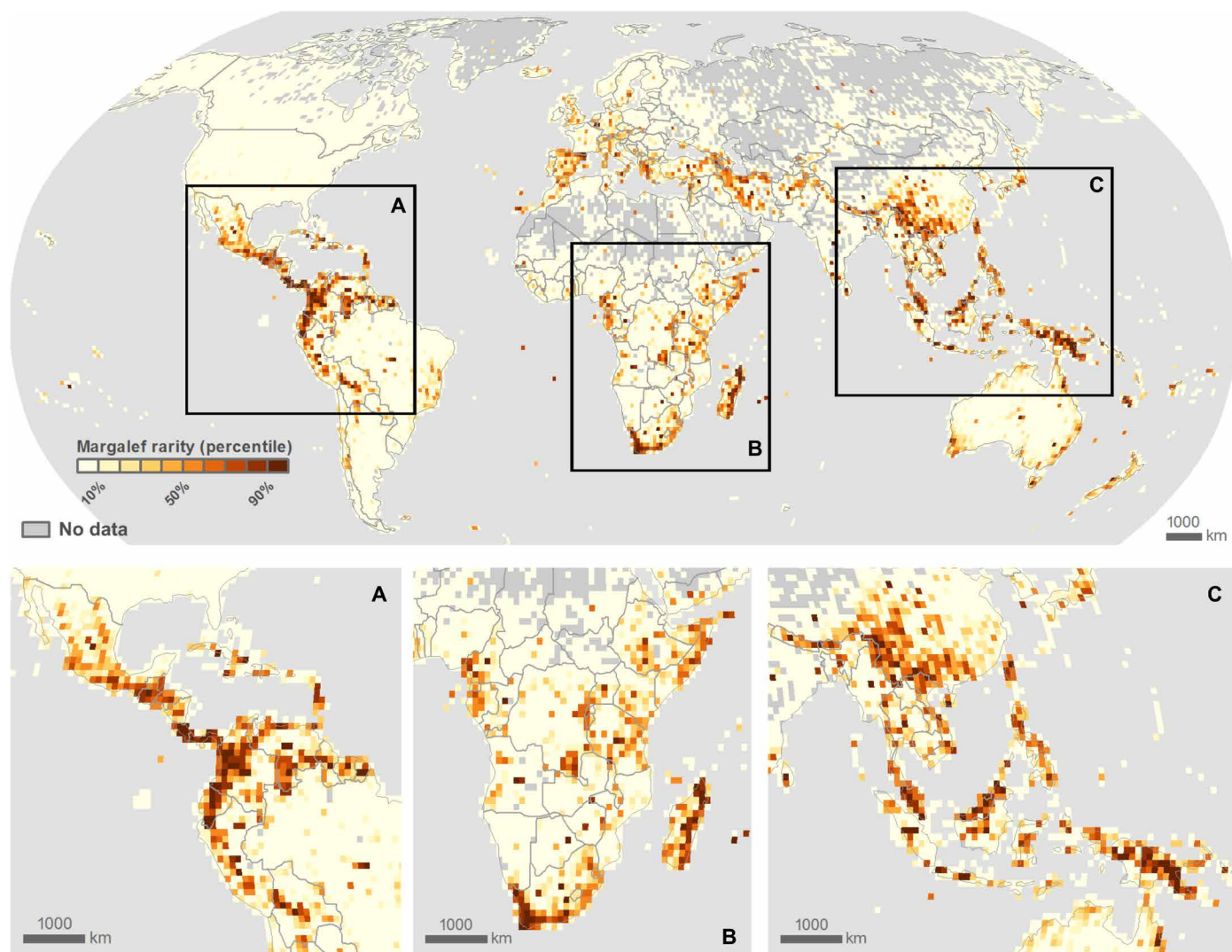


Fig. 4. Where are rare species distributed geographically? Plotting the geographic coordinates for all the observations for species with three observations or fewer at a coarse, 1° resolution reveals several patterns. The sampling background is shown (grey cells are areas with no georeferenced botanical sampling records, while yellow cells indicate regions with observation records but no rare species). Colored cells correspond to areas with rare species (species with three observations or fewer) rarified to the sampling intensity using the Margalef index (see the Supplementary Materials). Areas with a proportionally high number of rare species are dark brown (“hotspots of rarity”), while areas with relatively low numbers of rare species are yellow to orange. Areas with a high number of rare species tend to be clustered in a small number of locations including mountainous tropical and subtropical regions including New Guinea, Indonesia, southwestern China, Madagascar, the Andes (in Ecuador, Columbia, and Peru), Central America (Costa Rica and Panama), and southern Mexico. In addition, several notable temperate zone locations including the Fynbos in South Africa and southwest Australia, Northern Iran/Georgia/Turkey, and the Iberian Peninsula.

This is because areas with concentrations of rare species have previously been characterized by relatively more stable climates, but under the predicted climate change under RCP8.5, they will now experience velocities as high as the rest of the globe (see fig. S5).

Predicted changes of rarity indices

With the previously calibrated OLS and SAR full models, we made predictions of rarity indices under future projected climate. These showed worldwide decreases in rarity indices (Fig. 6), with the southern Andes and Southeast Asia predicted to experience the largest decreases. These decreases were likely due to the accelerated future climate velocities under RCP8.5, which are two orders of magnitude higher

than those experienced from LGM [~21 thousand years (ka) ago] to the present day (see fig. S5). Note, however, that future velocities are estimated over a shorter time frame, which will tend to produce higher estimates.

DISCUSSION

Our dataset represents the most comprehensive assembly of global plant diversity data to date, comprising both plots and herbarium specimens, from far more sources than previously available. Large quantities of primary biodiversity data have still not been mobilized, and those data that are available are subject to various forms of

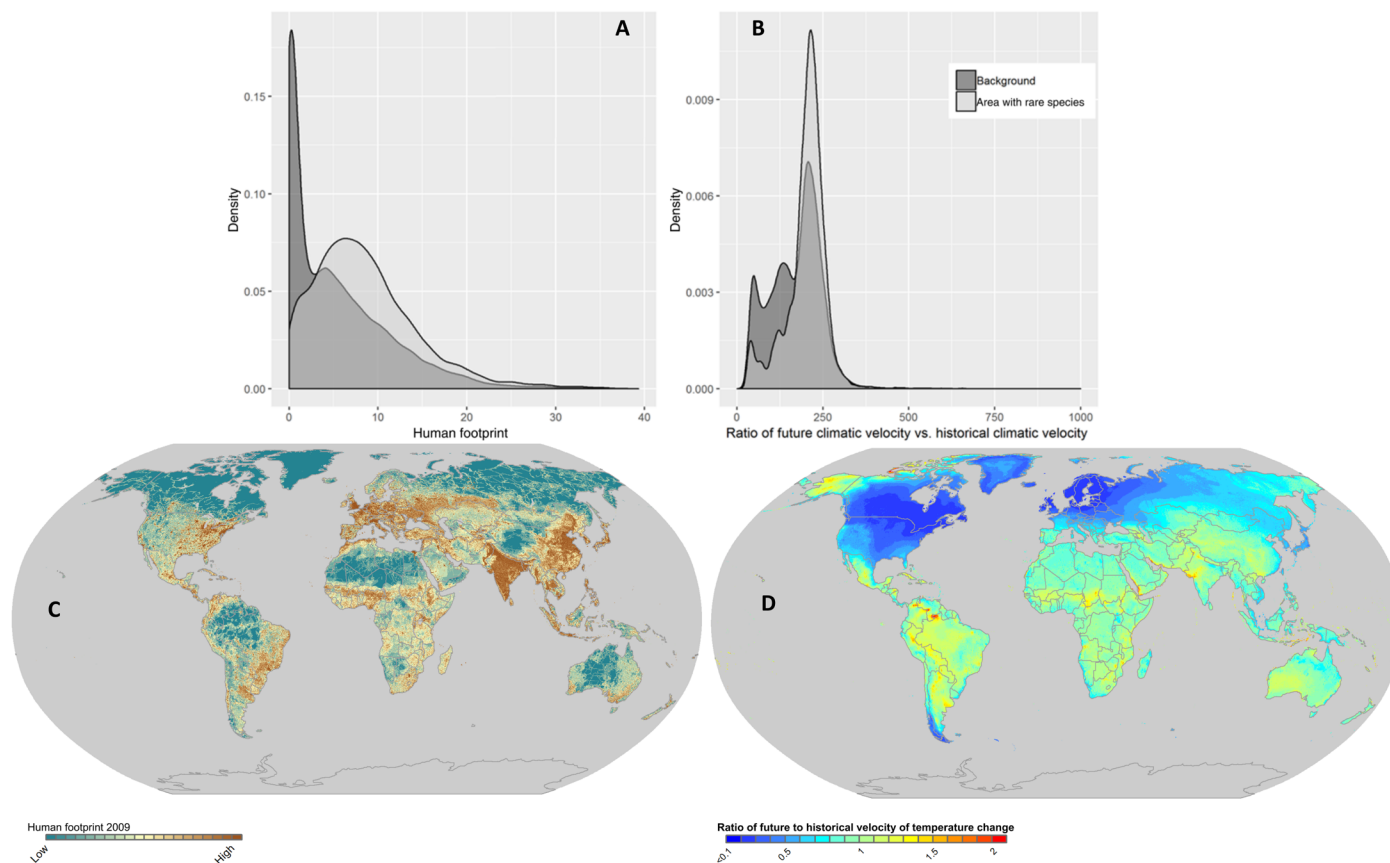


Fig. 5. Regions that currently have high numbers of rare species are also characterized by higher human impact and will experience faster rates of future climate change. (A) Density plot of human footprint index in areas with rare species (light gray) and the global map (dark gray). Areas with rare species have, on average, human footprint values of 8.5 ± 5.8 , which is ~ 1.6 times higher ($P < 0.001$, Wilcoxon test) human impact than on the globe on average (5.2 ± 5.8). (B) Density plot of the ratio of future climate (temperature) velocity versus historical climate velocity. On average, areas with rare species will experience $\sim 200 (\pm 58)$ times greater rates of temperature velocity than those same areas experienced historically and will experience ~ 1.2 times greater ($P < 0.001$, Wilcoxon test) rates of temperature velocity change than the globe will experience on average (170 ± 77). (C) Global variation in the human footprint index. Areas with high human footprint are in brown. Areas with low human footprint are dark green. (D) Global map of the ratio between future (baseline climate to late century, 1960–1990 to 2060–2080, under RCP8.5) and historical rates of temperature change [LGM to baseline climate (~ 21 ka ago to 1960–1990)]. Future temperatures will increase across the globe. However, in comparison with historical rates of climate change, some areas will experience relatively faster (ratio values greater than 1; yellow to red values) or slower (ratio values less than 1; green to blue values) rates of change. Note that many of the regions of rarity hotspots are found in regions that will be experiencing relatively faster rates of climate change compared to historical rates of change.

collection bias (17, 18). Thus, it is possible that the patterns we observe may change with additional data. However, comparison between plot and all observation gSADs (Fig. 2 and Table 1) indicates that both sampling methods yield similar results. Furthermore, the notable efforts we made in data cleaning and curation assure that our analyses represent, by far, the best window yet into global commonness and rarity in plants.

Our results indicate that hotspots of plant biodiversity largely reflect the accumulation of very rare species. Assessing the predictions of the Unified Theory of Neutral Biogeography [UTNB; (25)] for the distribution of commonness and rarity across species enables us to reveal likely drivers of rarity. The UTNB assumes that species overlap in their niches and are equivalent in their rates of speciation, extinction, and dispersal (25). It implies that biodiversity arises at random, as the abundance of each species follows a random walk so that the distribution of abundances across species is given by a dynamic equilibrium of speciation and extinction. Our results show

that $\beta \approx -1.4$, indicating that the proportion of plant species that are rare is higher than expected from neutral processes. Given that rare species are orders of magnitude more likely to go extinct (4, 5) than more abundant species, this begs the question: Why do we observe a larger proportion of observed rare species than expected from neutral theory?

Our analyses (tables S1 to S8 and figs. S1 to S5) suggest two primary reasons. First, current hotspots of rare species (Fig. 4) likely reflect areas with lowered risk of historical extinction. Rare species are often found in geographic localities that have had more stable climates that have likely lowered the probability of extinction [see (4, 5)]. Models that include relative climate stability better explain both the locations of hotspots of rarity and the shape of rarity distributions. The finding that climate (in)stability is important in non-neutral models has important real-world implications for ecology and conservation.

Second, rare species are spatially clumped in ways that support mechanisms for generating and maintaining rare species articulated

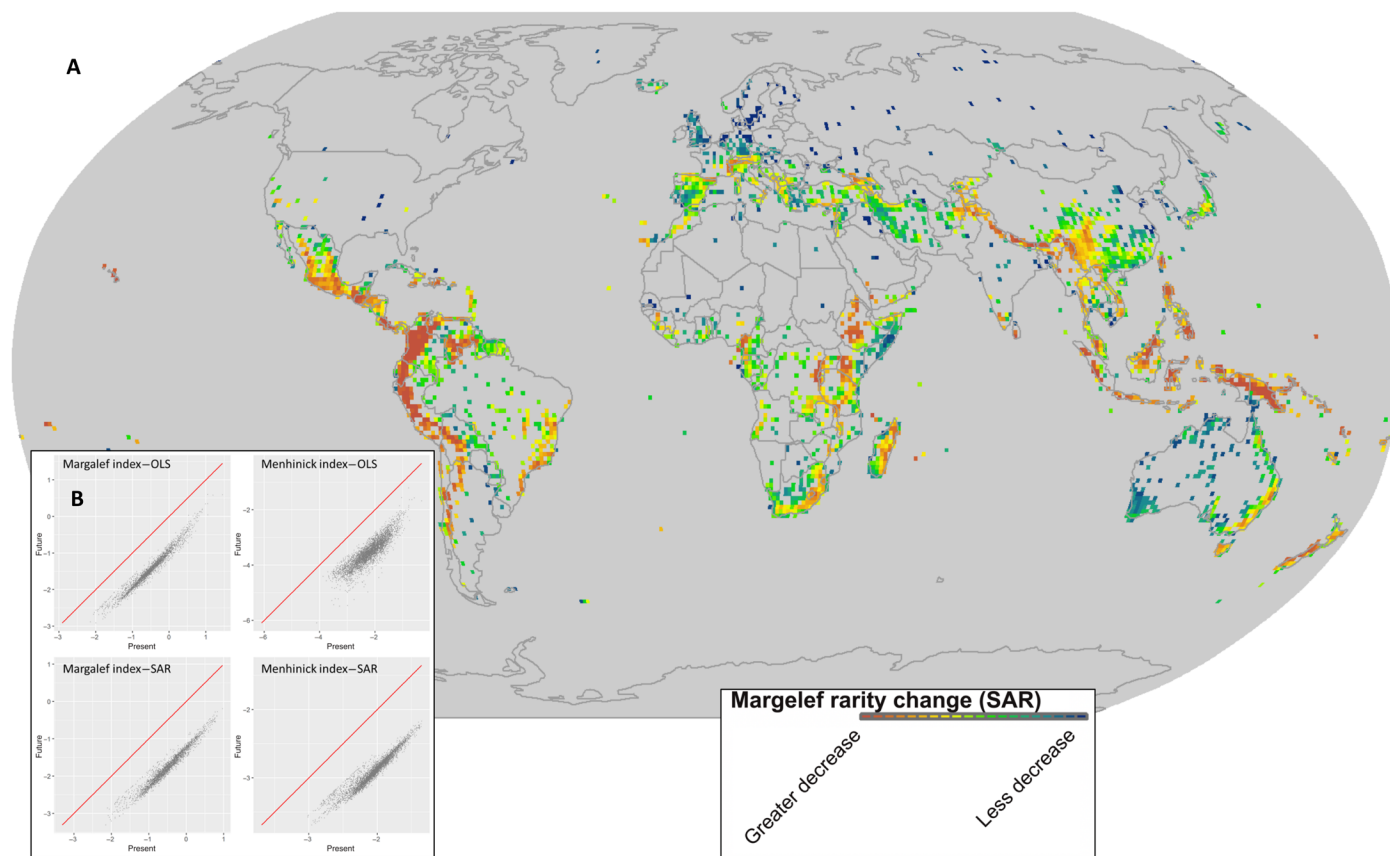


Fig. 6. What will happen to rare species diversity with climate change? (A) The predicted change in Margalef SAR rarity index under climate change from the autoregressive models (SAR). The rarity indices are log-transformed. Large decreases in climate suitability for rare species are in red to orange, whereas smaller reductions in climate suitability are given in green to blue colors. Note the large decreases in climate suitability for rare species in the Andes and Mesoamerica, African highlands, New Guinea, southwestern China, Indonesia, Nepal, and New Zealand. **(B)** The diagonal 1:1 line (red) represents situations of no difference between the predicted current and future rarity index from SAR and OLS models. All points in the scatter plot are below the diagonal line, indicating a reduction of rare species diversity across all the areas where they currently occur.

by early theorists, who proposed roles for mountains and climate stability in influencing both rates of speciation and dispersal. In 1964, Simpson (49) hypothesized that “Small population ranges and numerous barriers against the spread and sympatry of related populations would therefore tend to increase density of species.” Janzen’s 1967 (50) “Why mountain passes are higher in the tropics” extended Simpson’s hypothesis to predict that mountainous regions in the tropics will harbor proportionally more rare species than temperate mountains or even topographically uniform tropical regions due to less variability in climate. Our findings of disproportionate numbers of rare species in mainly tropical mountains and more isolated regions support these ideas. More recent studies have also documented the importance of tropical mountains as harbors of biodiversity (51, 52), which supports our findings.

Our results have important implications for conservation in the face of climate change and other human impacts. If ~36% of species are rare and threatened (Figs. 5 and 6), then ~158,000 plant species are at risk of extinction. Although not all primary biodiversity data have been digitized, it is still remarkable that ~36% of all plant species known are only documented a very small number of times. In addition, our analyses show that rapid rates of current human impact

and projected future climate change appear to disproportionately affect regions that harbor most of these rare species (Fig. 5), whereas the rare species likely have been in relatively more stable climates through their evolutionary history.

Ultimately, rare species, by definition, are more prone to reductions in population size and extinction and should be high priorities for conservation (4, 5). Our results suggest that redoubling global efforts to conserve rare species is needed and that we have a closing window to do so. The tools to ensure that these rare species are maintained are area-based conservation and solutions to climate change (53). The Convention on Biological Diversity should recognize these areas as critical to conserving all life on Earth and important focal areas for expansion of conserved areas after 2020 (54). The climate convention seeks to avoid extinctions due to the exceedance of species’ natural ability to adapt to climate change, making these areas with high numbers of rare species and very high future-to-historic velocities of climate change yet another reason the world should move quickly to curb greenhouse gas emissions (55). Joint climate and biodiversity efforts should be made to ensure that these numerous but little-known species, living in unusual climatic circumstances, persist into the future.

MATERIALS AND METHODS

Competing different hypothesized gSADs

As we described in the Supplementary Materials, we fit several additional hypothesized univariate distributions to the gSAD using the following proposed biological and statistical distributions. Most theories produced SADs that were so similar to each other that it was difficult to distinguish them given the noisy data and the fact that the differences were most pronounced in the tails, which were, by definition, infrequently observed (40). In Table 1, we provided several different goodness-of-fit measures. Each emphasizes different aspects of fit (chi-square on log-binned data emphasizes the fit of each statistical distribution to rare species, calculating an r^2 on the predicted versus empirical cumulative distribution function); cumulative distribution function [describes the probability that a random variable, X , drawn from $f(x)$ is $\leq x$] emphasizes the abundances with the most species (usually intermediate abundances), while likelihood emphasizes avoidance of extreme outliers. As previously noted, it is common for different measures of fit to select different SAD theories as providing the best fit to a single dataset (32). As a result, any claim of a superior fit must be robust by being superior on multiple measures.

Rarity indices

Because the sampling intensity for plants across the globe is not uniform, we assessed the rarified species diversity. For each 1° grid cell, we calculated the total number of observations or samples, N , as well as the total number of observed rare species, S ; for mapping rarity across the globe, we focused on the rarist species - those species having three observation records or fewer. We calculated two separate rarified diversity measures for each 1° grid cell:

1) Margalef diversity (S_{Margalef}), which assumes that species richness increases with sampling intensity N and, in particular, increases nonlinearly and approximately logarithmically with N .

$$S_{\text{Margalef}} = (S - 1) / \ln N \quad (3)$$

2) Menhinick diversity ($S_{\text{Menhinick}}$). In a similar vein, the Menhinick diversity measure assumes that species richness also increases nonlinearly with sampling intensity, N , but according to a square root function

$$S_{\text{Menhinick}} = S / \sqrt{N} \quad (4)$$

As the Menhenick index assumes a square root rarefaction function and the Margalef assumes a logarithmic rarefaction function, they represent both a more liberal and more conservative estimate of higher estimates of richness, respectively. Comparing both measures of S_{Margalef} and $S_{\text{Menhinick}}$ revealed similar spatial maps, indicating that both measures result in identical conclusions.

Methods for regression models

As described in the Supplementary Materials, we conducted OLS linear regression models to analyze the relationship between environmental variables and rarity index. We included three groups of environmental variables that portray present climate (annual mean temperature, annual precipitation, temperature seasonality, and precipitation seasonality), stability of climate (temperature velocity and precipitation velocity), and topology (elevation and heterogeneity of elevation). We also calculated the SD of elevations within

each one by 1° window and considered this as the heterogeneity of elevation. We performed log transformation of rarity index, temperature and precipitation velocity, elevation, and heterogeneity of elevation to get normally distributed residuals in the regression models. We standardized all variables to zero mean and 1 SD to make the regression coefficients comparable. With 4571 records (for Menhinick index, or 2940 for Margalef index) we conducted OLS linear regression models to explore the bivariate relationship between rarity index and each environmental variable.

We also constructed multiple regression models using each group of variables (present climate, stability of climate, and topology) and using all variables (full model). We conducted multiple regression models through exhaustive model selection based on AIC values using all environmental predictors. Last, to account for spatial autocorrelation in climate data, we performed Moran's I test and SAR models for all the OLS models mentioned above.

Climate change and future predicted changes in rarity indices

With the previously calibrated full models (OLS and SAR models), we made predictions of rarity indices under future projected climate. We used the full models, as they outperformed individual models or subgroup models and had comparable performances with the exhaustively selected model. We obtained future climatic variables from WorldClim (<http://www.worldclim.com/CMIP5v1>) (56). We used the future climate in 2070 constructed by the Community Climate System Model (CCSM4) under RCP8.5 scenario, which has comparatively high greenhouse gas emissions (57). To match the resolution of the rarity map, we sampled the environmental variables (annual temperature, annual precipitation, temperature seasonality, and precipitation seasonality) to 1° cells. We further calculated the temperature and precipitation velocity between present and future following (46). The two topological variables (elevation and heterogeneity of elevation) were kept the same as the present. After making the predictions, we compared the differences between predicted rarity indices under present and future climate.

Rarity and climate velocity

Using data sources and methods described above in regression model methods, we derived velocity of temperature change and velocity of precipitation change over the following periods: LGM to baseline climate (~21 ka ago to 1960–1990) and baseline climate to late century (1960–1990 to 2060–2080) (www.worldclim.org/paleo-climate1) under RCP8.5 (see Supplementary Document). Velocity was calculated using the neighborhood statistic approach originally described by Sandel *et al.* (46); see also (58). We note that our calculation of velocity of historical climate change and future climate change must be interpreted with caution, as they were calculated over different time intervals (46). We compared velocity values at locations where (i) there are rare species observations and (ii) there are no rare species observations and to (iii) background sampled locations. This comparison was conducted for both historical change since LGM and projected future change.

Rarity and the human footprint

We downloaded global human footprint data (48) and resampled to the resolution of the rarity map. We extracted the values of human footprint where rare species exist (i.e., 1° by 1° spatial windows where one or more rare species are observed) and compared the

mean of those values with that of the global human footprint map using the Wilcoxon test.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/5/11/eaaz0414/DC1>

Supplementary Document

Table S1. As in Table 1 but for specimen data found within 1° proximity to each plot.

Table S2. As with Table 2 but for specimens near plots.

Table S3. Summary statics of OLS linear regression models and SAR models for predicting the Menhinick rarity index.

Table S4. Summary statics of OLS linear regression models for predicting the Menhinick rarity index.

Table S5. Summary statics of SAR models for predicting the Menhinick rarity index.

Table S6. Summary statics of OLS linear regression models and SAR models for predicting the Margalef rarity index.

Table S7. Summary statics of OLS linear regression models for predicting the Margalef rarity index.

Table S8. Summary statics of SAR models for predicting the Margalef rarity index.

Fig. S1. Sampling density for different data types in BIEN.

Fig. S2. Scatter plots showing the relationships between bivariate relationship between Menhinick rarity index and environmental variables.

Fig. S3. Scatter plots showing the relationships between bivariate relationship between Margalef rarity index and environmental variables.

Fig. S4. Predicted changes of Margalef rarity index using either the OLS or the SAR models.

Fig. S5. Historical and future global temperature velocities.

References (59–108)

REFERENCES AND NOTES

1. F. W. Preston, The canonical distribution of commonness and rarity: Part I. *Ecology* **43**, 185–215 (1962).
2. I. Hanski, Dynamics of regional distribution: The core and satellite species hypothesis. *Oikos* **38**, 210–221 (1982).
3. C. Darwin, *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (John Murray, 1859), 502 pp.
4. J. M. Diamond, "Normal" extinctions of isolated populations, in *Extinctions*, M. H. Nitecki, Ed. (University of Chicago Press, 1984), pp. 191–246.
5. S. L. Pimm, H. L. Jones, J. Diamond, On the risk of extinction. *Am. Nat.* **132**, 757–785 (1988).
6. S. P. Hubbell, R. B. Foster, in *Conservation Biology: The Science of Scarcity and Diversity*, M. E. Soulé, Ed. (Sinauer Associates, 1986), pp. 205–231.
7. M. E. Soulé, *Conservation Biology: The Science of Scarcity and Diversity* (Sinauer Associates, 1986).
8. H. Ter Steege, N. C. A. Pitman, T. J. Killeen, W. F. Laurance, C. A. Peres, J. E. Guevara, R. P. Salomão, C. V. Castilho, I. L. Amaral, F. D. de Almeida Matos, L. de Souza Coelho, W. E. Magnusson, O. L. Phillips, D. de Andrade Lima Filho, M. de Jesus Veiga Carim, M. V. Irupe, M. P. Martins, J.-F. Molino, D. Sabatier, F. Wittmann, D. C. López, J. R. da Silva Guimarães, A. M. Mendoza, P. N. Vargas, A. G. Manzatto, N. F. C. Reis, J. Terborgh, K. R. Casula, J. C. Montero, T. R. Feldpausch, E. N. H. Coronado, A. J. D. Montoya, C. E. Zartman, B. Mostacedo, R. Vasquez, R. L. Assis, M. B. Medeiros, M. F. Simon, A. Andrade, J. L. Camargo, S. G. W. Laurance, H. E. M. Nascimento, B. S. Marimon, B.-H. Marimon Jr., F. Costa, N. Targhetta, I. C. G. Vieira, R. Brienen, H. Castellanos, J. F. Duivenvoorden, H. F. Mogollón, M. T. F. Piedade, G. A. C. Aymard, J. A. Comiskey, G. Damasco, N. Dávila, R. García-Villacorta, P. R. S. Diaz, A. Vicentini, T. Emilio, C. Levis, J. Schietti, P. Souza, A. Alonso, F. Dallmeier, L. V. Ferreira, D. Neill, A. Araujo-Murakami, L. Arroyo, F. A. Carvalho, F. C. Souza, D. D. do Amaral, R. Gribel, B. G. Luiz, M. P. Pansonato, E. Venticinque, P. Fine, M. Toledo, C. Baraloto, C. Cerón, J. Engel, T. W. Henkel, E. M. Jimenez, P. Maas, M. C. P. Mora, P. Petronelli, J. D. C. Revilla, M. Silveira, J. Stropp, R. Thomas-Caesar, T. R. Baker, D. Daly, M. R. Paredes, N. F. da Silva, A. Fuentes, P. M. Jørgensen, J. Schöngart, M. R. Silman, N. C. Arboleda, B. B. L. Cintra, F. C. Valverde, A. Di Fiore, J. F. Phillips, T. R. van Andel, P. von Hildebrand, E. M. Barbosa, L. C. de Matos Bonates, D. de Castro, E. de Sousa Farias, T. Gonzales, J.-L. Guillaumet, R. Hoffman, Y. Malhi, I. P. de Andrade Miranda, A. Prieto, A. Rudas, A. R. Ruschell, N. Silva, C. I. A. Vela, V. A. Vos, E. L. Zent, S. Zent, A. Cano, M. T. Nascimento, A. A. Oliveira, H. Ramirez-Angulo, J. F. Ramos, R. Sierra, M. Tirado, M. N. U. Medina, G. van der Heijden, E. V. Torre, C. Vriesendorp, O. Wang, K. R. Young, C. Baider, H. Balslev, N. de Castro, W. Farfan-Rios, C. Ferreira, C. Mendoza, I. Mesones, A. Torres-Lezama, L. E. U. Giraldo, D. Villarroel, R. Zagt, M. N. Alexiades, K. Garcia-Cabrera, L. Hernandez, I. Huamantupa-Chuquimaco, W. Milliken, W. P. Cuenca, S. Pansini, D. Pualetto, F. R. Arevalo, A. F. Sampaio, E. H. Valderrama Sandoval, L. V. Gamarra, Estimating the global conservation status of more than 15,000 Amazonian tree species. *Sci. Adv.* **1**, e1500936 (2015).
9. A. Zizka, H. Ter Steege, M. d. C. R. Pessoa, A. Antonelli, Finding needles in the haystack: Where to look for rare species in the American tropics. *Ecography* **41**, 321–330 (2018).
10. H. Ter Steege, R. W. Vaessen, D. Cárdenas-López, D. Sabatier, A. Antonelli, S. M. de Oliveira, N. C. A. Pitman, P. M. Jørgensen, R. P. Salomão, The discovery of the Amazonian tree flora with an updated checklist of all known tree taxa. *Sci. Rep.* **6**, 29549 (2016).
11. J. H. Brown, D. W. Mehlman, G. C. Stevens, Spatial variation in abundance. *Ecology* **76**, 2028–2043 (1995).
12. S. R. Abades, P. A. Marquet, Finite size scaling in the local abundances of geographic populations. *Biol. Res.* **44**, 107–112 (2011).
13. F. He, K. J. Gaston, Occupancy-abundance relationships and sampling scales. *Ecography* **23**, 503–511 (2000).
14. J. H. Brown, *Macroecology* (University of Chicago Press, 1995).
15. B. J. McGill, Does Mother Nature really prefer rare species or are log-left-skewed SADs a sampling artefact? *Ecol. Lett.* **6**, 766–773 (2003).
16. C. N. Jenkins, S. L. Pimm, L. N. Joppa, Global patterns of terrestrial vertebrate diversity and conservation. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E2602–E2610 (2013).
17. B. H. Daru, D. S. Park, R. B. Primack, C. G. Willis, D. S. Barrington, T. J. S. Whitfield, T. G. Seidler, P. W. Sweeney, D. R. Foster, A. M. Ellison, C. C. Davis, Widespread sampling biases in herbaria revealed from large-scale digitization. *New Phytol.* **217**, 939–955 (2018).
18. C. Meyer, P. Weigelt, H. Kref, Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecol. Lett.* **19**, 992–1006 (2016).
19. B. J. Enquist, R. Condit, R. K. Peet, M. Schildhauer, B. M. Thiers, "Cyber in frastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity" (PeerJ Preprints, 2016); <https://peerj.com/preprints/2615.pdf>.
20. C. Thomas, Biodiversity databases spread, prompting unification call. *Science* **324**, 1632–1633 (2009).
21. D. Rabinowitz, Seven forms of rarity, in *The Biological Aspects of Rare Plant Conservation*, H. Synge, Ed. (Wiley, 1981), pp. 205–217.
22. J. Yu, F. S. Dobson, Seven forms of rarity in mammals. *J. Biogeogr.* **27**, 131–139 (2000).
23. IUCN, The IUCN Red List of Threatened Species (2019); www.iucnredlist.org.
24. A. E. Magurran, P. A. Henderson, Explaining the excess of rare species in natural species abundance distributions. *Nature* **422**, 714–716 (2003).
25. S. P. Hubbell, *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, 2001).
26. R. A. Chisholm, S. W. Pacala, Niche and neutral models predict asymptotically equivalent species abundance distributions in high-diversity ecological communities. *Proc. Nat. Acad. Sci. U.S.A.* **107**, 15821–15825 (2010).
27. R. A. Fisher, A. S. Corbet, C. B. Williams, The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **12**, 42–58 (1943).
28. S. Pueyo, Diversity: Between neutrality and structure. *Oikos* **112**, 392–405 (2006).
29. E. P. White, B. J. Enquist, J. L. Green, On estimating the exponent of power-law frequency distributions. *Ecology* **89**, 905–912 (2008).
30. T. Zillio, R. Condit, The impact of neutrality, niche differentiation and species input on diversity and abundance distributions. *Methods* **116**, 931–940 (2007).
31. J. Aitchison, J. A. C. Brown, *The Lognormal Distribution with Special Reference to Its Uses in Economics* (Cambridge Univ. Press, 1957); <http://cds.cern.ch/record/109563>.
32. B. McGill, Strong and weak tests of macroecological theory. *Oikos* **102**, 679–685 (2003).
33. M. Williamson, K. J. Gaston, The lognormal distribution is not an appropriate null hypothesis for the species-abundance distribution. *J. Anim. Ecol.* **74**, 409–422 (2005).
34. R. May, Patterns of species abundance and diversity, in *Ecology and Evolution of Communities*, M. L. Cody, J. M. Diamond, Eds. (Harvard Univ. Press, 1975), pp. 81–120.
35. R. H. MacArthur, On the relative abundance of species. *Am. Nat.* **94**, 25–36 (1960).
36. E. C. Pielou, *Ecological Diversity* (John Wiley & Sons, 1975).
37. S. Engen, R. Lande, Population dynamic models generating the log normal species abundance distribution. *Math. Biosci.* **132**, 169–183 (1996).
38. E. C. Pielou, *Mathematical Ecology* (Wiley, 1977).
39. F. P. Schoenberg, R. Peng, J. Woods, On the distribution of wildfire sizes. *Environmetrics* **14**, 583–592 (2003).
40. B. J. McGill, R. S. Etienne, J. S. Gray, D. Alonso, M. J. Anderson, H. K. Bence, M. Dornelas, B. J. Enquist, J. L. Green, F. He, A. H. Hurlbert, A. E. Magurran, P. A. Marquet, B. A. Maurer, A. Ostling, C. U. Soykan, K. I. Ugland, E. P. White, Species abundance distributions: Moving beyond single prediction theories to integration within an ecological framework. *Ecol. Lett.* **10**, 995–1015 (2007).
41. S. R. Connolly, T. P. Hughes, D. R. Bellwood, R. H. Karlson, Community structure of corals and reef fishes at multiple scales. *Science* **309**, 1363–1365 (2005).
42. S. L. Pimm, P. H. Raven, The fate of the World's plants. *Trends Ecol. Evol.* **32**, 317–320 (2017).
43. E. N. Lughadha, R. Govaerts, I. Belyaeva, N. Black, H. Lindon, R. Allkin, R. E. Magill, N. Nicolson, Counting counts: Revised estimates of numbers of accepted species of flowering plants, seed plants, vascular plants and land plants with a review of other recent estimates. *Phytotaxa* **272**, 82–88 (2016).

44. N. C. A. Pitman, J. W. Terborgh, M. R. Silman, P. Núñez V, D. A. Neill, C. E. Cerón, W. A. Palacios, M. Aulestia, Dominance and distribution of tree species in upper Amazonian Terra Firme Forests. *Ecology* **82**, 2101–2117 (2001).
45. N. Myers, R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, J. Kent, Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).
46. B. Sandel, L. Arge, B. Dalsgaard, R. G. Davies, K. J. Gaston, W. J. Sutherland, J.-C. Svenning, The influence of late Quaternary climate-change velocity on species endemism. *Science* **334**, 660–664 (2011).
47. B. Sandel, A.-C. Monnet, R. Govaerts, M. Vorontsova, Late quaternary climate stability and the origins and future of global grass endemism. *Ann. Bot.* **119**, 279–288 (2017).
48. O. Venter, E. W. Sanderson, A. Magrath, J. R. Allan, J. Beher, K. R. Jones, H. P. Possingham, W. F. Laurance, P. Wood, B. M. Fekete, M. A. Levy, J. E. M. Watson, Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. *Nat. Commun.* **7**, 12558 (2016).
49. G. G. Simpson, Species density of North American recent mammals. *Syst. Zool.* **13**, 57–73 (1964).
50. D. H. Janzen, Why mountain passes are higher in the tropics. *Am. Nat.* **101**, 233–249 (1967).
51. C. Rahbek, M. K. Borregaard, A. Antonelli, R. K. Colwell, B. G. Holt, D. Nogués-Bravo, C. M. Ø. Rasmussen, K. Richardson, M. T. Rosing, R. J. Whittaker, J. Fjeldså, Building mountain biodiversity: Geological and evolutionary processes. *Science* **365**, 1114–1119 (2019).
52. C. Rahbek, M. K. Borregaard, R. K. Colwell, B. Dalsgaard, B. G. Holt, N. Morueta-Holme, D. Nogués-Bravo, R. J. Whittaker, J. Fjeldså, Humboldt's enigma: What causes global patterns of mountain biodiversity? *Science* **365**, 1108–1113 (2019).
53. L. Hannah, G. Midgley, S. Andelman, M. Araújo, G. Hughes, E. Martinez-Meyer, R. Pearson, P. Williams, Protected area needs in a changing climate. *Front. Ecol. Environ.* **5**, 131–138 (2007).
54. E. Dinerstein, C. Vynne, E. Sala, A. R. Joshi, S. Fernando, T. E. Lovejoy, J. Mayorga, D. Olson, G. P. Asner, J. E. M. Baillie, N. D. Burgess, K. Burkart, R. F. Noss, Y. P. Zhang, A. Baccini, T. Birch, N. Hahn, L. N. Joppa, E. Wikramanayake, A Global Deal For Nature: Guiding principles, milestones, and targets. *Sci. Adv.* **5**, eaaw2869 (2019).
55. Intergovernmental Panel on Climate Change, *Climate Change 2014—Impacts, Adaptation and Vulnerability: Part A: Global and Sectoral Aspects: Volume 1, Global and Sectoral Aspects: Working Group II Contribution to the IPCC Fifth Assessment Report* (Cambridge Univ. Press, 2014).
56. R. J. Hijmans, S. E. Cameron, J. L. Parra, P. G. Jones, A. Jarvis, Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
57. K. Riahi, S. Rao, V. Krey, C. Cho, V. Chirkov, G. Fischer, G. Kindermann, N. Nakicenovic, P. Rafaj, RCP 8.5—A scenario of comparatively high greenhouse gas emissions. *Clim. Change* **109**, 33 (2011).
58. S. R. Loarie, P. B. Duffy, H. Hamilton, G. P. Asner, C. B. Field, D. D. Ackerly, The velocity of climate change. *Nature* **462**, 1052–1055 (2009).
59. J. Wiecezorek, D. Bloom, R. Guralnick, S. Blum, M. Döring, R. Giovanni, T. Robertson, D. Vieglais, Darwin Core: An evolving community-developed biodiversity data standard. *PLOS ONE* **7**, e29715 (2012).
60. B. Boyle, N. Hopkins, Z. Lu, J. A. Raygoza Garay, D. Mozzherin, T. Rees, N. Matasci, M. L. Narro, W. H. Piel, S. J. McKay, S. Lowry, C. Freeland, R. K. Peet, B. J. Enquist, The taxonomic name resolution service: An online tool for automated standardization of plant names. *BMC Bioinf.* **14**, 16 (2013).
61. B. S. Maitner, B. Boyle, N. Casler, R. Condit, J. Donoghue, S. M. Durán, D. Guaderrama, C. E. Hinchliff, P. M. Jørgensen, N. J. B. Kraft, B. McGill, C. Merow, N. Morueta-Holme, R. K. Peet, B. Sandel, M. Schildhauer, S. A. Smith, J.-C. Svenning, B. Thiers, C. Violle, S. Wiser, B. J. Enquist, The BIEN R package: A tool to access the Botanical Information and Ecology Network (BIEN) database. *Methods Ecol. Evol.* **9**, 373–379 (2018).
62. G. R. Goldsmith, N. Morueta-Holme, B. Sandel, E. D. Fitz, S. D. Fitz, B. Boyle, N. Casler, K. Engemann, P. M. Jørgensen, N. J. B. Kraft, B. McGill, R. K. Peet, W. H. Piel, N. Spencer, J. C. Svenning, B. M. Thiers, C. Violle, S. K. Wiser, B. J. Enquist, *Plant-O-Matic*: A dynamic and mobile guide to all plants of the Americas. *Methods Ecol. Evol.* **7**, 960–965 (2016).
63. I. R. McFadden, B. Sandel, C. Tsirogianis, N. Morueta-Holme, J.-C. Svenning, B. J. Enquist, N. J. B. Kraft, Temperature shapes opposing latitudinal gradients of plant taxonomic and phylogenetic β diversity. *Ecol. Lett.* **22**, 1126–1135 (2019).
64. E. Fegraus, Tropical Ecology Assessment and Monitoring Network (TEAM Network). *Biodivers. Ecol.* **4**, 287 (2012).
65. R. K. Peet, M. T. Lee, M. D. Jennings, D. Faber-Langendoen, VegBank: A permanent, open-access archive for vegetation plot data. *Biodivers. Ecol.* **4**, 233–241 (2012).
66. S. J. DeWalt, G. Bourdy, L. R. Chávez de Michel, C. Quenevo, Ethnobotany of the Tacana: Quantitative inventories of two permanent plots of Northwestern Bolivia. *Econ. Bot.* **53**, 237–260 (1999).
67. Forest Inventory and Analysis National Program (2013); www.fia.fs.fed.us/.
68. S. K. Wiser, P. J. Bellingham, L. E. Burrows, Managing biodiversity information: Development of New Zealand's National Vegetation Survey databank. *N. Z. J. Ecol.* **25**, 1–17 (2001).
69. K. J. Anderson-Teixeira, S. J. Davies, A. C. Bennett, E. B. Gonzalez-Akre, H. C. Muller-Landau, S. J. Wright, K. Abu Salim, A. M. Almeyda Zambrano, A. Alonso, J. L. Baltzer, Y. Basset, N. A. Bourg, E. N. Broadbent, W. Y. Brockelman, S. Bunyavejchewin, D. F. Burslem, N. Butt, M. Cao, D. Cardenas, G. B. Chuyong, K. Clay, S. Cordell, H. S. Dattaraja, X. Deng, M. Detto, X. Du, A. Duque, D. L. Erikson, C. E. Ewango, G. A. Fischer, C. Fletcher, R. B. Foster, C. P. Giardina, G. S. Gilbert, N. Gunatilleke, S. Gunatilleke, Z. Hao, W. W. Hargrove, T. B. Hart, B. C. Hau, F. He, F. M. Hoffman, R. W. Howe, S. P. Hubbell, F. M. Inman-Narahari, P. A. Jansen, M. Jiang, D. J. Johnson, M. Kanzaki, A. R. Kassim, D. Kenfack, S. Kibet, M. F. Kinnaird, L. Korte, K. Kral, J. Kumar, A. J. Larson, Y. Li, X. Li, S. Liu, S. K. Lum, J. A. Lutz, K. Ma, D. M. Maddalena, J. R. Makana, Y. Malhi, T. Marthews, R. Mat Serudin, S. M. McMahon, W. J. McShea, H. R. Memiaghe, X. Mi, T. Mizuno, M. Morecroft, J. A. Myers, V. Novotny, A. A. de Oliveira, P. S. Ong, D. A. Orwig, R. Ostertag, J. den Ouden, G. G. Parker, R. P. Phillips, L. Sack, M. N. Sainge, W. Sang, K. Sri-Ngernyuan, R. Sukumar, I. F. Sun, W. Sungpalee, H. S. Suresh, S. Tan, S. C. Thomas, D. W. Thomas, J. Thompson, B. L. Turner, M. Uriarte, R. Valencia, M. I. Vallejo, A. Vicentini, T. Vrška, X. Wang, X. Wang, G. Weiblen, A. Wolf, H. Xu, S. Yap, J. Zimmerman, CTF5-ForestGEO: A worldwide network monitoring forests in an era of global change. *Glob. Chang. Biol.* **21**, 528–549 (2015).
70. B. Enquist, B. Boyle, SALVIAS—The SALVIAS vegetation inventory database. *Biodivers. Ecol.* **4**, 288 (2012).
71. GBIF.org, GBIF Occurrence Download; 10.15468/dl.yubndf.
72. G. Dauby, R. Zaiss, A. Blach-Overgaard, L. Catarino, T. Damen, V. Deblauwe, S. Dessein, J. Dransfield, V. Droissart, M. C. Duarte, H. Engledow, G. Fadeur, R. Figueira, R. E. Gereau, O. J. Hardy, D. J. Harris, J. de Heij, S. Janssens, Y. Klomberg, A. C. Ley, B. A. Mackinder, P. Meerts, J. L. van de Poel, B. Sonké, M. S. M. Sosef, T. Stévant, P. Stoffelen, J.-C. Svenning, P. Sepulchre, X. van der Burgt, J. J. Wieringa, T. L. P. Couvreur, RAINBIO: A mega-database of tropical African vascular plants distributions. *PhytoKeys*, 1–18 (2016).
73. G. Arellano, V. Cala, A. Fuentes, L. Cayola, P. M. Jørgensen, M. J. Macia, A standard protocol for woody plant inventories and soil characterisation using temporary 0.1-Ha plots in tropical forests. *J. Trop. For. Sci.* **28**, 508–516 (2016).
74. B. Crandall-Stotler, R. E. Stotler, D. G. Long, Morphology and classification of the Marchantiophyta, in *Bryophyte Biology*, A. J. Shaw, Ed. (Cambridge Univ. Press, 2008), pp. 1–54.
75. E. M. Gifford, A. S. Foster, *Morphology and Evolution of Vascular Plants* (W. H. Freeman, 1989).
76. B. Goffinet, W. Buck, Systematics of the Bryophyta (mosses): From molecules to a revised classification. *Monogr. Syst. Botany Missouri Bot. Garden* **98**, 205–239 (2004).
77. P. H. Raven, R. F. Evert, S. E. Eichhorn, *Biology of Plants* (Macmillan, 2005).
78. H. Bischler, R. M. Schuster, The Hepaticae and Anthocerotae of North America. *Bryologist* **96**, 281 (1993).
79. B. J. McGill, Species abundance distributions, in *Biological Diversity: Frontiers in Measurement & Assessment*, A. E. Magurran, B. J. McGill, Eds. (Oxford Univ. Press, 2011), pp. 105–122.
80. B. A. Maurer, B. J. McGill, Measurement of species diversity, in *Biological Diversity: Frontiers in Measurement and Assessment*, A. E. Magurran, B. J. McGill, Eds. (Oxford Univ. Press, 2011), pp. 55–65.
81. M. Evans, N. Hastings, B. Peacock, *Mathematical Distributions* (1993).
82. R. H. MacArthur, On the relative abundance of bird species. *Proc. Natl. Acad. Sci. U.S.A.* **43**, 293–295 (1957).
83. I. Motomura, A statistical treatment of associations. *Jap. J. Zool.* **44**, 379–383 (1932).
84. F. W. Preston, The canonical distribution of commonness and rarity: Part II. *Ecology* **43**, 410–432 (1962).
85. B. Mandelbrot, Information theory and psycholinguistics: A theory of word frequencies, in *Readings in Mathematical Social Sciences*, P. F. Lazarsfeld, N. W. Henry, Eds. (Science Research Associates Inc., 1966).
86. A. J. Kerkhoff, B. J. Enquist, Multiplicative by nature: Why logarithmic transformation is necessary in allometry. *J. Theor. Biol.* **257**, 519–521 (2009).
87. F. Galton, The geometric mean, in vital and social statistics. *Proc. R. Soc. London* **29**, 365–367 (1879).
88. P. D. Gingerich, Arithmetic or geometric normality of biological variation: An empirical test of theory. *J. Theor. Biol.* **204**, 201–221 (2000).
89. D. McAlister, XIII. The law of the geometric mean. *Proc. R. Soc. London* **29**, 367–376 (1879).
90. S. R. Hallyo, A theoretical framework for abundance distributions in complex systems. *Complex. Int.* **6**, 1–16 (1998).
91. F. He, D. Tang, Estimating the niche preemption parameter of the geometric series. *Acta Oecol.* **33**, 105–107 (2008).
92. A. E. Magurran, *Ecological Diversity and Its Measurement* (Princeton Univ. Press, 1988).
93. B. C. McCarthy, A. E. Magurran, Measuring biological diversity. *J. Torrey Bot. Soc.* **131**, 277 (2004).
94. J. Izsák, Some practical aspects of fitting and testing the Zipf-Mandelbrot model: A short essay. *Scientometrics* **67**, 107–120 (2006).
95. M. G. Bulmer, On fitting the Poisson lognormal distribution to species-abundance data. *Biometrics* **30**, 101–110 (1974).
96. E. R. Ziegel, M. Evans, N. Hastings, B. Peacock, Statistical distributions. *Technometrics* **36**, 229 (1994).
97. B. J. McGill, A test of the unified neutral theory of biodiversity. *Nature* **422**, 881–885 (2003).

98. R. Margaleff, Temporal succession and spatial heterogeneity in phytoplankton, in *Perspectives in Marine Biology*, A. A. Buzzati-Traverso, Ed. (University of California, 1958), pp. 323–349.
99. E. F. Menhinick, A comparison of some species-individuals diversity indices applied to samples of field insects. *Ecology* **45**, 859–861 (1964).
100. G. Feng, Z. Ma, B. Sandel, L. Mao, S. Normand, A. Ordóñez, J.-C. Svenning, Species and phylogenetic endemism in angiosperm trees across the Northern Hemisphere are jointly shaped by modern climate and glacial-interglacial climate change. *Glob. Ecol. Biogeogr.* **28**, 1393–1402 (2019).
101. G. Feng, L. Mao, B. Sandel, N. G. Swenson, J.-C. Svenning, High plant endemism in China is partially linked to reduced glacial-interglacial climate change. *J. Biogeogr.* **43**, 145–154 (2016).
102. H. Kreft, W. Jetz, Global patterns and determinants of vascular plant diversity. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5925–5930 (2007).
103. D. S. Park, O. H. Razafindratsima, Anthropogenic threats can have cascading homogenizing effects on the phylogenetic and functional diversity of tropical ecosystems. *Ecography* **42**, 148–161 (2019).
104. W. D. Kissling, G. Carl, Spatial autocorrelation and the selection of simultaneous autoregressive models. *Glob. Ecol. Biogeogr.* **17**, 59–71 (2008).
105. R. Development Core Team, R: A language and environment for statistical computing (2019); www.r-project.org/.
106. R. J. Hijmans, J. van Etten, M. Mattiuzzi, M. Sumner, J. A. Greenberg, O. P. Lamigueiro, A. Bevan, E. B. Racine, A. Shortridge, Raster package in R (2013); www.r-spatial.org/raster/RasterPackage.pdf.
107. V. Calcagno, C. de Mazancourt, glmulti: An R package for easy automated model selection with (generalized) linear models. *J. Stat. Softw.* **34**, 1–29 (2010).
108. R. Bivand, L. Anselin, O. Berke, A. Bernat, M. Carvalho, Y. Chun, C. F. Dormann, S. Dray, R. Halbersma, N. Lewin-Koh, H. Ono, P. Peres-Neto, M. Tiefseldorf, D. Yu, spdep: Spatial dependence: Weighting schemes, statistics and models (2011); <http://ftp.auckland.ac.nz/software/CRAN/src/contrib/Descriptions/spdep.html>.

Acknowledgments: This work was conducted as a part of the BIEN Working Group, 2008–2012. We thank all the data contributors and numerous herbaria who have contributed their data to various data compiling organizations (see the Supplementary Materials) for the invaluable data and support provided to BIEN. We thank the New York Botanical Garden; Missouri Botanical Garden; Utrecht Herbarium; the UNC Herbarium; and GBIF, REMIB, and SpeciesLink. The staff at CyVerse provided critical computational assistance. We thank the more than 50 scientists who participated in our various BIEN working group and subgroup meetings since 2008 including B. Blonder, K. Engemann, E. Fegraus, J. Cavender-Bares, B. Dobrin, K. Gendler, R. Jorgensen, G. Lopez-Gonzalez, L. Zhenyuan, S. McKay, O. Phillips, J. Pickering, N. Swenson, C. Vriesendorp, and K. Woods, who participated in a working group meeting, and D. Ackerly, E. Garnier, R. Guralnick, W. Jetz, J. Macklin, N. Matasci, S. Ramteke, and A. Zanne who participated in subgroup meetings. We also acknowledge the critical support of the University of Arizona high-performance computing resources via the Research Data Center as well as iPlant and CyVerse support from R. Jorgensen, S. Goff, N. Matasci, N. Merchant, M. Narrow, and R. Walls. Furthermore, the long-term vision, encouragement, and computational support of F. Davis, S. Hampton, M. Jones, N. Outin, and the ever-helpful staff at NCEAS were critical for the completion of this first stage of the BIEN working group. Special thanks to K. Koenig for cartographic support. We acknowledge the herbaria that contributed data to this work: A, AAH, AAS, AAU, ABH, ACAD, ACOR, AD, AFS, AK, AKPM, ALCB, ALTA, ALU, AMD, AMES, AMNH, AMO, ANGU, ANSM, ANSP, AQP, ARAN, ARIZ, AS, ASDM, ASU, AUT, AV, AWH, B, BA, BAA, BAB, BABY, BACP, BAF, BAFC, BAI, BAJ, BAL, BARC, BAS, BBB, BBS, BC, BCMEX, BCN, BCRU, BERE, BESA, BG, BH, BHCB, BIO, BISH, BLA, BM, BOCH, BOL, BOLV, BONN, BOON, BOTU, BOUM, BPI, BR, BREM, BRI, BRIT, BRLU, BRM, BSB, BUT, C, CALI, CAN, CANB, CANU, CAS, CATA, CATIE, CAY, CBM, CDA, CDBI, CEN, CEPEC, CESS, CGE, CGMS, CHAM, CHAPA, CHAS, CHR, HSC, CIB, CICY, CIIDIR, CIMI, CLEM, CLEMS, CLF, CMX, CMMEX, CNPO, CNS, COA, COAH, COCA, CODAGEM, COFC, COL, COLO, CONC, CORD, CP, CPAP, CPUN, CR, CRAI, CRP, CS, CSU, CSUSB, CTES, CTESN, CU, CUV, CUZ, CVRD, DAO, DAV, DBG, DBN, DES, DLF, DNA, DPU, DR, DS, DSM, DUKE, DUSS, E, EA, EAC, EAN, EBUM, ECON, EIF, EIU, EMMA, ENCB, ER, ERA, ESA, ETH, F, FAA, FAU, FAUC, FB, FCME, FCO, FCQ, FEN, FHO, FI, FLAS, FLOR, FM, FR, FRU, FSU, FTG, FUEL, FULD, FURB, G, GAT, GB, GDA, GENT, GES, GH, GI, GLM, GMDRC, GMNHJ, GOET, GRA, GUA, GZU, H, HA, HAC, HAL, HAM, HAMAB, HAO, HAS, HASU, HB, HBG, HBR, HCIB, HEID, HGM, HIB, HIP, HNT, HO, HPL, HRCB, HRP, HSC, HSS, HU, HUA, HUAA, HUAL, HUAZ, HUCP, HUEFS, HUEN, HUFU, HUJ, HUSA, HUT, HXHB, HYO, IAA, IAC, IAN, IB, IBGE, IBK, IBSC, IBUG, ICLE, ICESI, ICN, IEA, IEB, ILL, IMSSM, INB, INEGI, INIF, INM, INPA, IPA, IPRN, IRVC, ISC, ISKW, ISL, ISTC, ISU, IZAC, IZTA, JACA, JBAG, JBG, JCT, JE, JEPS, JOTR, JROH, JUA, JYV, K, KIEL, KMN, KMNH, KOELN, KOR, KPM, KSC, KSTC, KSU, KUT, KU, KUN, KYO, L, LA, LAGU, LBG, LD, LE, LEB, LIL, LINC, LINN, LISE, LISI, LISU, LL, LMS, LOJA, LOMA, LP, LPAG, LPB, LPD, LPS, LSU, LSUM, LTB, LTR, LW, LYJB, LZ, M, MA, MACF, MAF, MAK, MARS, MARY, MASS, MB, MBK, MBM, MBML, MCNS, MEL, MELU, MEN, MERL, MEXU,

MFA, MFU, MG, MGC, MICH, MIL, MIN, MISSA, MJG, MMMN, MNHM, MNHN, MO, MOL, MOR, MPN, MPU, MPUC, MSB, MSC, MSUN, MT, MTMG, MU, MUB, MUR, MVFA, MVFO, MVJB, MVM, MW, MY, N, NA, NAC, NAS, NCU, NE, NH, NHM, NHMC, NHT, NLH, NM, NMB, MNML, NMR, NMSU, NSPM, NSW, NT, NU, NUM, NY, NZFRI, O, OBI, ODU, OS, OSA, OSC, OSH, OULU, OWU, OXF, P, PACA, PAMP, PAR, PASA, PDD, PE, PEL, PERTH, PEUF, PFC, PGM, PH, PKDC, PLAT, PMA, POM, PORT, PR, PRC, PRE, PSU, PY, QCA, QCNE, QFA, QM, QRS, QUE, R, RAS, RB, RBR, REG, RELC, RFA, RIQC, RM, RNG, RSA, RYU, S, SACT, SALA, SAM, SAN, SANT, SAPS, SASK, SAV, SBBG, SBT, SCFS, SD, SDSU, SEL, SEV, SF, SFV, SGO, SI, SIU, SJRP, SJSU, SLPM, SMDB, SMF, SNM, SOM, SP, SPF, SPSF, SQF, SRFA, STL, STU, SUU, SVG, TAES, TAI, TAIF, TALL, TAM, TAMU, TAN, TASH, TEF, TENN, TEPB, TEX, TFC, TI, TKPM, TNS, TO, TOYA, TRA, TRH, TROM, TRT, TRTE, TU, TUB, U, UADY, UAM, UAMIZ, UB, UBC, UC, UCMM, UCR, UCS, UCSB, UCSC, UEC, UESC, UFG, UFMA, UFMT, UFP, UFRJ, UFRN, UFS, UGDA, UH, UI, UJAT, ULM, ULS, UME, UMO, UNA, UNB, UNCC, UNEX, UNITEC, UNL, UNM, UNR, UNSL, UPBC, UPEI, UPNA, UPS, US, USAS, USF, USJ, USM, USNC, USP, USZ, UT, UTC, UTEP, UU, UVIC, UWO, V, VAL, VALD, VDB, VEN, VIT, VMSL, VT, W, WAG, WAT, WELT, WFU, WI, WIN, WIS, WMNH, WOLL, WS, WTU, WU, XAL, YAMA, Z, ZMT, ZSS, and ZT. **Funding:** The BIEN working group was supported by the National Center for Ecological Analysis and Synthesis, a center funded by NSF EF-0553768 at the University of California, Santa Barbara and the State of California. Additional support for the BIEN working group was provided by iPlant/CyVerse via NSF DBI-0735191. B.J.E., B.J.M., and C.M. were supported by NSF ABI-1565118. B.J.E. and C.M. were supported by NSF HDR-1934790. B.J.E., L.H., R.T.C., G.M., W.F., and P.R.R. were supported by the Global Environment Facility SPARC project grant (GEF-5810). N.M.-H. was supported by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 746334 and acknowledges the Danish National Research Foundation for support to the Center for Macroecology, Evolution and Climate (grant no. DNRF96). C.V. was supported by a Marie Curie International Outgoing Fellowship within the 7th European Community Framework Program (DiversiTraits project no. 221060) and by the European Research Council (ERC) Starting Grant Project (grant ERC-StG-2014-639706-CONSTRAINTS). B.J.E., B.M., N.J.B.K., C.V., and B.J.M. acknowledge the FREE group funded by the synthesis center CESAB of the French Foundation for Research on Biodiversity (FRB) and EDF. J.-C.S. and B.J.E. acknowledge support from the Center for Informatics Research on Complexity in Ecology (CIRCE), funded by the Aarhus University Research Foundation under the AU Ideas program. J.-C.S. also considers this work a contribution to his VILLUM Investigator project "Biodiversity dynamics in a changing world" funded by VILLUM FONDEN (grant 16549). X.F., D.S.P., and E.A.N. were supported by the University of Arizona Bridging Biodiversity and Conservation Science program. C.M. acknowledges funding from NSF Grant DBI-1913673. S.K.W. acknowledges funding from the Strategic Science Investment Fund to Crown Research Institutes from the New Zealand Ministry of Business, Innovation and Employment. I.S. was supported by the Charles University (UNCE 204069). T.L.P.C., G.D., and J.J.W. acknowledge the French FRB and the Provence-Alpes-Côte d'Azur région (PACA) through the Centre for Synthesis and Analysis of Biodiversity (CESAB) data program, as part of the RAINBIO research project. **Author contributions:** B.J.E. and B.J.M. designed the study. B.B., X.F., B.M., and D.S.P. integrated and cleaned data. P.M.J., B.M.T., T.L.P.C., G.D., D.M.N., A.T.O.-F., R.K.P., J.M.S.-D., J.J.W., W.F., and S.K.W. contributed data. B.J.E., B.J.M., X.F., B.M., B.B., and P.R.R. performed the analyses. B.J.E., B.J.M., J.M.S.-D., L.H., P.M.J., B.M.T., X.F., B.B., B.S., E.A.N., P.M.J., P.R.R., B.M.T., J.R.B., R.T.C., T.L.P.C., J.C.D., J.C.L., P.A.M., C.M., G.M., N.M.-H., N.J.B.K., D.S.P., R.K.P., M.P., J.M.S.-D., B.S., M.S., I.S. C.V., S.K.W., and J.-C.S. helped interpret and analyze results. All authors helped collect and assemble data. B.J.E. wrote the first draft of the manuscript, and all authors contributed to revisions. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. All data and code used in this study are available via GitHub, https://github.com/EnquistLab/BIEN_Rarity. The MATLAB code used for fitting gSADs are available from B.J.M. upon request. The public version of the BIEN database is accessible via the BIEN R package, <https://cran.r-project.org/web/packages/BIEN/index.html>. Additional data related to this paper may be requested from the authors.

Submitted 8 August 2019
 Accepted 4 November 2019
 Published 27 November 2019
 10.1126/sciadv.aaz0414

Citation: B. J. Enquist, X. Feng, B. Boyle, B. Maitner, E. A. Newman, P. M. Jørgensen, P. R. Roehrdanz, B. M. Thiers, J. R. Burger, R. T. Corlett, T. L. P. Couvreur, G. Dauby, J. C. Donoghue, W. Foden, J. C. Lovett, P. A. Marquet, C. Merow, G. Midgley, N. Morueta-Holme, D. M. Neves, A. T. Oliveira-Filho, N. J. B. Kraft, D. S. Park, R. K. Peet, M. Pillet, J. M. Serra-Diaz, B. Sandel, M. Schildhauer, I. Šimová, C. Violle, J. J. Wieringa, S. K. Wiser, L. Hannah, J.-C. Svenning, B. J. McGill, The commonness of rarity: Global and future distribution of rarity across land plants. *Sci. Adv.* **5**, eaaz0414 (2019).